Semi-Automatic Classification of Cementitious Materials using Scanning Electron Microscope Images

Lucas Drumetz, Mauro Dalla Mura, Samuel Meulenyzer, Sébastien Lombard, Jocelyn Chanussot

Semi-Automatic Classification of Cementitious Materials using Scanning Electron Microscope Images

Lucas Drumetz,[a] Mauro Dalla Mura,[a] Samuel Meulenyzer,[b] Sébastien Lombard[b], Jocelyn Chanussot[a]

[a]Grenoble Institute of Technology, GIPSA-lab, Image and Signal Dept., 11 Rue des Mathématiques, Saint-Martin d'Hères, France, 38402

[b]Lafarge Centre de Recherche, 95 Rue du Montmurier, Saint-Quentin-Fallavier, France, 38070

Abstract. Segmentation and classification are prolific research topics in the image processing community, which have been more and more used in the context of analysis of cementitious materials, on images acquired with Scanning Electron Microscopes (SEM) . Indeed, there is a need to be able to detect and to quantify the materials present in a cement paste in order to follow the chemical reactions occurring in the material even days after the solidification. In this paper, we propose a new approach for segmentation and classification of cementitious materials based on the denoising of the data with the Block Matching 3D (BM3D) algorithm, Binary Partition Tree (BPT) segmentation, Support Vector Machines (SVM) classification, and the interactivity with the user. The BPT provides a hierarchical representation of the spatial regions of the data, allowing a segmentation to be selected among the admissible partitions of the image. SVMs are used to obtain a classification map of the image. This approach combines state-of-the-art image processing tools with the interactivity with the user to allow a better segmentation to be performed, or to help the classifier discriminate the classes better. We show that the proposed approach outperforms a previous method on synthetic data and several real datasets coming from cement samples, both qualitatively with visual examination and quantitatively with the comparison of experimental results with theoretical ones.

Keywords: SEM image,image segmentation, classification, image denoising, interactivity.

Address all correspondence to: Lucas Drumetz, GIPSA-lab, Image and Signal Dept., 11 Rue des Mathématiques, Saint-Martin d'Hères, France, 38402; Tel: +33 (0)4 76 82 63 13 ; Fax: + 33 (0)4 76 57 47 90;
E-mail: lucas.drumetz@gipsa-lab.fr

1 Introduction

Digital Image Processing techniques have been extensively used in the past few decades for the analysis of microscope images.[1] Specifically, the analysis of Scanning Electron Microscopy (SEM) images has been performed for various applications such as biology, physics and materials science. Indeed, SEM allows the acquisition of high resolution images of the surface of a material using different physical phenomena occuring after the sample of interest is excited by an electron beam.[2] SEM images and image processing techniques can be used for the analysis of building materials. One application is the analysis of the microstructure of cement paste in order to quantify the proportions of some phases of interest in the sample (here and in all this paper, the word phase must

be understood as a thermodynamic state where the physical properties of the cement are constant). The ultimate goal is to be able to use this information to monitor the chemical hydration reactions taking place in cementitious materials, by inferring the hydration degree of the components. These reactions continue for months after the cement paste is formed. Monitoring them opens the path to a better understanding of cement behavior over time. The cement industry is also willing to use the monitoring of the chemical reactions in order to produce more durable and resistant materials, and also to test new more environmentally-friendly cement compositions, as the cement industry is an important source of CO_2 emissions.

The identification of the components of the cement in SEM images can be rather accurately performed visually by a specialist, but there is a need to automatize the procedure, or at least reduce the contribution of an expert to a minimum. That is why image segmentation and classification techniques have been used in order to be able to cluster the data in several similar regions corresponding to specific cement phases, or to identify the components present in cement samples in a supervised fashion. Frequently, only the Backscattered Electron image is used to discriminate between the phases.[3] However, the information present in this type of data often is insufficient to classify accurately all phases since their gray levels ranges may overlap. Besides, some specific compounds can make the task more difficult to solve: for instance, the so-called fly ashes, which are a pozzolanic material resulting from the calcination of coal, have been more and more used in the composition of cementitious materials due to their positive impact on the energy and natural resources expenses required to produce cement. This type of mostly circular compounds can be of various sizes, sometimes very small with respect to the resolution of SEM images (resulting in sometimes only a few pixels for a single micronized fly ash), which makes the processing even more difficult. Besides, they can be shallow, resulting in shadowing effects on the acquired images

that are easy to interpret visually, but hard to process in an automatic way. All these aspects show the need for new efficient image processing techniques for SEM images of cementitious materials, specifically in the domains of image segmentation and classification.

In this paper, we propose an integrated processing pipeline for SEM image processing in order to obtain automatic segmentation and classification maps, and especially allow easy and efficient interaction with the user for both steps. The main contribution lies in the interactivity enabled by the proposed framework, and in the assembly of the existing tools in a processing chain suited for SEM image processing. The proposed method was first outlined in a conference paper,[4] with much less detail on the process and the validation.

The rest of this paper is organized as follows. Section 2 introduces the data we will work with and the problem to solve. Section 3 briefly reviews the existing methods to tackle the problem. Section 4 presents the solution we propose. Section 5 presents the results obtained using the proposed approach in comparison with other existing techniques. Finally, section 6 provides some concluding remarks and proposes a few leads for future work on the topic.

2 Presentation of the data and the problem

Here, we will describe and formalize the problem we are faced with and the data acquired and used for this study. As stated in the introduction, the data is acquired with a Scanning Electron Microscope[2] on a prepared cement test sample. The preparation of the sample, a quite delicate task, will not be detailed here, but a description of the process can be found in other works.[5,6] For this application, as it has been the case in previous works, we will use the SEM Backscattered Electron (BSE) image and also the X-ray diffraction images at carefully chosen excitation energies. Each of these images taken individually is a gray-scale image. In the following, we will refer to one

of these gray-scale images as a "band", using the vocabulary of multispectral and hyperspectral imaging. BSE images are high spatial resolution images (354 nm here) whose gray levels are related to the density of the material at the pixel location, while X-ray images, once the excitation energies are well tuned, can account for the presence of the corresponding chemical elements in the sample. The most common elements present in cement samples have been selected in the context of this study. It results in 14 X-ray bands for the following elements: Al, C, Ca, Cl, Fe, K, Mg, Mn, Na, O, P, S, Si and Ti. Note that the BSE image provides a much higher spatial resolution than the X-ray bands, and hence it can be seen as providing a better representation of the spatial arrangement of the structures in the image, whereas the bundle of X-ray bands provide spectral information, since they correspond to the same physical property observed at different excitation energies.[6] The resulting images can be seen together as one single data cube, with two dimensions accounting for the spatial coordinates, and the last one for the "spectral" coordinates (although the term is a bit misleading here since we have two different types of data, *i.e.* BSE and X-ray bands). An example of an image obtained after the acquisition is shown in Fig. 1. Additionally, we will have to deal with noise in the X-ray bands, as well as quantization effects occurring during the acquisition process for the scarcest elements of the samples. This aspect will have to be accounted for in the way we pre-process the data. We illustrate this quantization effect with an example given in Fig. 2.

Our problem is to be able to identify the composition of the cement, for example by being able to detect clusters of similar pixels, thus identifying meaningful regions. This is a segmentation problem and it can be addressed in an unsupervised way. In the end, we want to quantify each phase of the cement. To do so, we have to assign every pixel to a given pre-determined class corresponding to one of the components present in the cement sample. For our application, a

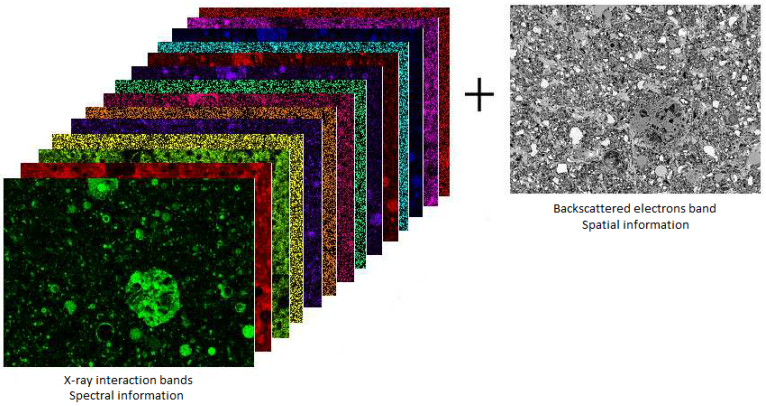

Figure 1 A SEM image of a cement sample. The X-ray bands were assigned a colormap for an easier identification of the corresponding chemical elements (the more saturated the color, the higher the value). The first X-ray image (green) accounts for the presence of Aluminium in the sample. The width of the images represents roughly $350\,\mu\mathrm{m}$.

training set of a priori labelled pixels is used, making the problem a supervised classification problem. The rationale of this approach is that it is quite fast to identify and label some pixels of the image according to the phase they belong to, but it would be much too tedious to do this for the whole image, let alone for batches of images of the same sample. On the other hand, training a classifier on one image will allow to classify multiple images of the same sample. We show an example of training samples in our case in Fig. 3. If we can obtain a classification map, we will be able to get the percentages of every phase in the sample. Finally, using samples of the same cement at different times, the monitoring of the chemical reactions in the cement paste becomes possible, for example by computing the hydration degree of a phase (the unreacted fraction of the phase, at a certain time after the solidification of the cement paste). In the next section, we will review some existing methods for segmentation and classification.

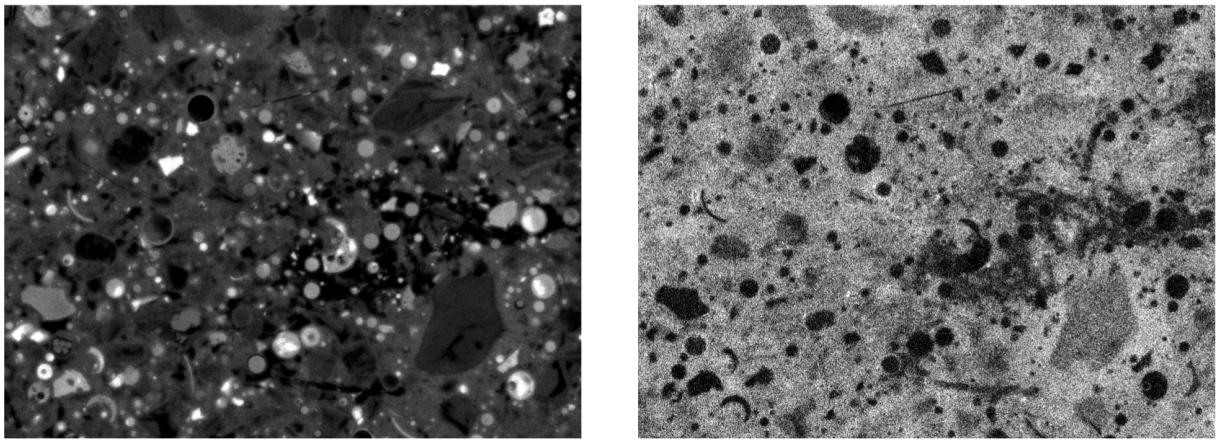

Figure 2 Two X-ray bands showing different configurations of noise and quantization in a 91 days-old cement sample. On the left is a noisy X-ray interaction band corresponding to an abundant chemical element (Si), and on the right is a heavily quantized band related to a scarce chemical element (Cl).

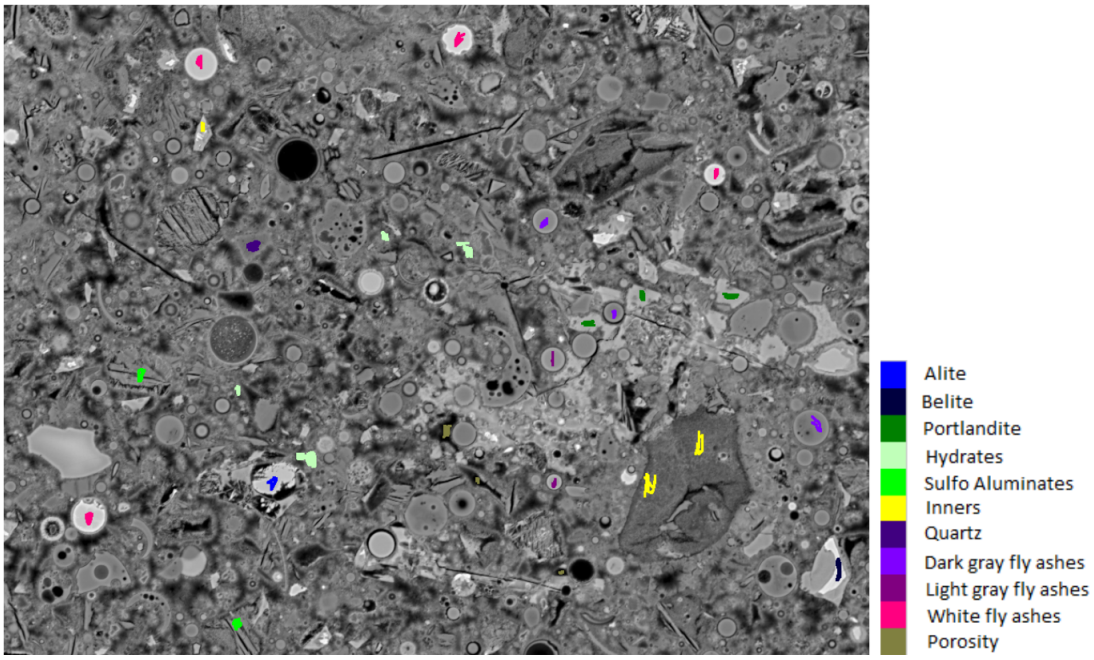

Figure 3 The training image consists in the BSE band with some colored pixels corresponding to the different classes.

3 Existing techniques

Here, we present some existing segmentation and classification algorithms used in the cement research community and also in the image processing community.

3.1 Image classification and segmentation for concrete and cement

Image analysis on SEM images is one of the techniques used to characterize the microstructure of a cement paste. As the final goal is classification, none of the following methods are specifically designed for segmentation, which is only a temporary step to allow the classification to be performed.

Most existing methods are based on thresholding or on the use of mathematical morphology operators to segment the images. Some methods rely only on the BSE image to discriminate some phases, such as porosity, which will be systematically associated to the low gray-levels in the BSE image. Sometimes, techniques are specifically designed to process a given phase. An example or a mathematical morphology-based algorithm aimed at detecting fly ashes and estimating their hydration degree was introduced by Haha et al.[3] These simple techniques can only be used for some of the phases that can be found on cement paste because the only use the information of the BSE image, which can provide too little information for some phases, such as fly ashes whose gray levels in the BSE bands overlap with calcium sulfoaluminates. As the histograms of the acquired bands of cement samples are often multimodal, some techniques perform a segmentation using multiple thresholds. For instance, another work[7] uses multiple thresholds the BSE image and the X-ray bands, and then uses a priori information on the cement phases (described as low, medium or high density for the BSE image, and low, medium or high concentration of the chemical elements of interest for the X-ray images). A similar approach for different types of cement is presented by

Stutzman.[8] In the work by Silva et al.,[9] a segmentation and hydration degree estimation method is proposed. It uses mathematical morphology operators to compute the gradient of the image. Then the segmentation itself is based on the well-known watershed algorithm.[10]

A more recent approach[6] is directly classification-oriented since it does not make use of any segmentation map. It uses the Support Vector Machines (SVM[11]) classification technique to make use of all the data provided by the BSE image and the X-ray bands. It is a supervised classification technique, in the sense that some pixels are manually assigned a label corresponding to the class they belong to thanks to the examination of the user. Then these pixels are used as training samples for the SVM. The features used are simply the original bands of the image. SVMs are originally binary classifiers which are well suited if the data are linearly separable in the feature space. However, even when they are not, the classifier is able to propose a linear border between the classes minimizing the misclassifications. In any case, the idea is to find the hyperplane that best separates the two classes defined by the training instances. Then the data are assigned to one of the classes depending on the half-space they belong to. This binary classifier can be easily extended to the multi-class case. However, if the data are not linearly separable in the feature space, we can use the so-called kernel trick to map the data into a higher dimensional space in which, hopefully, the data will be linearly separable. The resulting classification map is noisy since the SVM process is a pixel-wise classifier and as such, it does not take into account the spatial correlation between neighboring pixels. To tackle this, a spatial regularization technique uses Markov Random Fields (MRF[12]) to smooth the SVM classification map, though this regularization comes at the price of a loss of some spatial details in the edge pixels of the identified phases. Smaller components, such as micronized fly ashes which can be only be a few pixels wide tend to disappear after the regularization. This method, denoted here as SVM-MRF, will be used in section 5 for comparison with

the results of the proposed approach.

3.2 Image classification and segmentation in the image processing community

Image segmentation is still one of the most active research topics in image processing. Over the years, a plethora of methods have been developed to this end, be it for gray-scale, color or multivariate images. We will mention some of them in this section. Note that the list we give here is far from being exhaustive. Other than the threshold and watershed based-methods, there exist multiple variants of region-based algorithms. We can cite for example the region-growing or split-and-merge classes of algorithms.[13] We can also mention the mean-shift algorithm,[14] which has been successfully applied to SEM images.[15] Markov Random Fields have also been used for image segmentation purposes.[16]Graph-based segmentation methods, as well as fuzzy logic based segmentation have also been investigated[17,18] . Finally, a hierarchical segmentation framework was more recently proposed. The ill-posedness of the segmentation problem leads to the absence of an optimal segmentation for complex scenes. Based on this idea, the Binary Partition Tree (BPT) was proposed.[19] It provides a tree representation, in which every node represents a region of the image. The idea is to combine nodes corresponding to similar regions in a pair-wise mode. The merging of two regions will define the hierarchical relations of the tree. We show an example of the construction of a BPT and its pruning in Fig. 4. This approach provides a multiscale representation of the image, each level of the hierarchy corresponding to a possible segmentation, a partition of the image in meaningful regions, *i.e.* homogeneous with respect to some pre-defined criterion. On the lower levels of the hierarchy, the segmentation will be very fine (oversegmentation), and it will become coarser and coarser as the level of hierarchy increases. The lowest level correspond to a starting partition (it can be the whole image, in which case the initial regions are all the pixels

of the image, or an initial segmentation computed by other means). The nodes of the tree corresponding to the regions of the initial partition are called the leaves of the tree. The top levels will be extremely coarse (undersegmentation), and the root of the tree is a region whose support is the same as the whole image. For a practical application, however, a precise segmentation map has to be extracted from the BPT, *i.e.* a well chosen partition among the admissible ones given the tree structure. This step is called pruning the tree, since the goal is to keep only certain nodes of the tree to define a partition of the image. The pruning is completely application-dependent (segmentation, classification, object recognition...).[20] The most simple pruning strategy is a cross-cut at a certain height of the tree. The BPT is adapted to multi-band images, such as the ones we are dealing with. Sections 4.2.1 and 4.2.2 provide a more thorough description of this technique. It has been recently successfully used in the context of multivalued data in hyperspectral imaging.[21,22] Another hierarchical segmentation approach was developed allowing the preservation of a certain criterion in each region of the partition. The algorithm can be adapted for interactive video sequence segmentation,[23] using propagation of user defined markers.

Image segmentation is an ill-posed problem with no optimal solution, even for a specific application, making it hard to assess the results in a quantitative way. Nevertheless, in the last decade, some techniques to evaluate the quality of segmentation have surfaced, be it in a unsupervised way,[24] or by comparing the output of some algorithm to an available reference.[25]

As for classification techniques, then again, a wide range of methods have been developed. We will focus here only classical methods for supervised classification, in which some a priori information, *i.e.* training samples, are available. These training samples consist in a set of pixels whose class (referred to as label) is identified before the processing. Among these approaches are the nearest neighbor classifiers, decision trees[26] (not to be mistaken with the BPT, which can

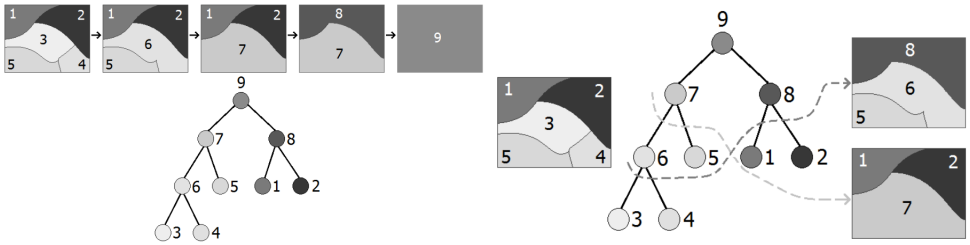

Figure 4 An example of the construction of a BPT (left). Starting from an initial partition (the leaves, here nodes 1 to 5), the most similar regions are merged, creating a new node in the tree, and a new partition. This process is iterated until only one region remains (the root of the tree). Two possible prunings of the BPT (right). We can only select nodes that form a partition of the initial image.

also be used for classification purposes using particular pruning strategies[21]), or Support Vector Machines (SVM). Refer to the book of Bishop[27] for a review. Note the the idea of hierarchy can also be considered for classification at different levels.[28] SVMs have also successfully been used in the hyperspectral imaging context.[29] They have also been coupled with Markov Random Fields (MRF) to provide a spatial regularization of the results of the SVM, that are always noisy since they are a pixel-wise procedure.[30] As stated before, MRFs have the disadvantage of destroying some information on the edges of the regions and on small objects. In order to spatially regularize the classification maps, SVMs can also be run on a segmentation of the image, such as a watershed segmentation or a superpixel generation algorithm.[31] Finally, sparse representation based algorithms have received more and more attention over the last decade.[32,33]

4 Proposed approach

In this section, we present in detail the approach we propose to perform both segmentation and classification of the cement multivalued images in an interactive way. Indeed, the user can intervene at some points of the processing to improve the results of the automatic segmentation and of the supervised classification. The flowchart of the proposed approach is given in Fig. 5. This section is divided in three parts: first we present the pre-processing step we perform, namely the

denoising of the X-ray bands. Then we introduce the proposed strategy to perform a segmentation of the multivariate cement images, using BPTs, and how we allow the user to improve it. Finally, we present the technique we use to get a classification map given some training samples, and once again we explain how the user is able to improve the results by adjusting the training set, with the help of the segmentation map computed during the previous step.

The contributions of the paper are the following:

- We propose an automatic processing chain (which combines several tools into a pipeline) suited for SEM image processing, especially in case of low SNR and quantization of the X-ray bands. This pipeline also provides all the necessary tools for user interaction.

- We lay out two strategies to improve the automatic results via user interaction: for segmentation by exploring the hierarchical representation given by the BPT, and for classification by introducing user decisions into the training set using a segmentation map.

- We introduce a novel pruning strategy for Binary Partition Tree (BPT) based image segmentation using discontinuities in the attributes of the image regions along a branch of the tree.

While many steps are involved in the process because of the challenging data used (very noisy, subject to quantization effects, multimodal and multispectral, spatially complex), most of them are fully automatic and help obtaining good automatic segmentation and classification maps. The user interaction allows to easily improve significantly the classification and segmentation results when necessary (as in this case with complex datasets), making the proposed method an global and flexible processing pipeline.

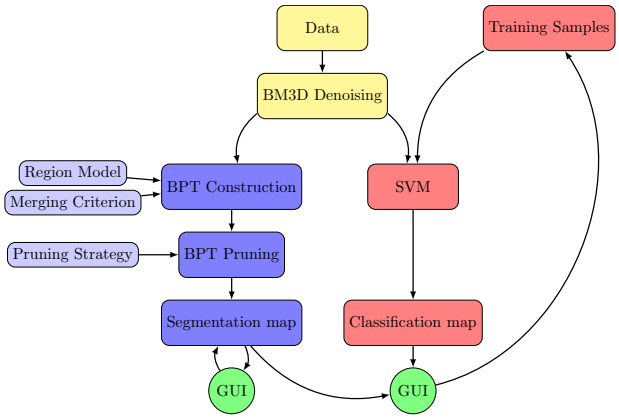

Figure 5 Flowchart of the proposed method. The blue boxes correspond to the segmentation part of the process, while the red ones correspond to the classification part. Green circles show the parts where the process interacts with the user, to improve the segmentation map or the training samples.

4.1 Data pre-processing

The first step of the proposed approach is the denoising of the data. Simple traditional approaches such as the median or mean filters fail to significantly denoise the data without altering the spatial structures of the images, especially when quantization effects occur. For our problem, we chose to use the Block Matching 3D (BM3D) algorithm.[34] It is based on the assumption that the noise is i.i.d. Gaussian with zero mean, and processes gray-level images by dividing them into blocks of size 8×8. Algthough the noise in X-ray images can have very specific properties, we chose to use this strategy because of the additional quantization effect of our images, which make denoising all the more difficult. The BM3D algorithm is a two step procedure: first a basic estimate of the denoised image is computed as follows: similar blocks are stacked in a 3D cube. The idea behind this block matching is to use the spatial correlation of the image, as well as the correlation between similar blocks. An example of the block matching procedure is shown in Fig. 6. Without going into details, the 3D stacked blocks are projected in a transformed domain using wavelets and Discrete Cosine transforms. In the transformed domain, the signal is sparsely represented in a few

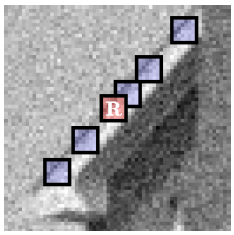

Figure 6 An example of the block matching procedure to find similar blocks in a gray-scale image.[34] A reference block (R) is shown in red, and the blue blocks (found here along an edge of a spatial structure of the image) were identified as similar.

coefficients while the noise coefficients are dense. By hard thresholding, the noise contribution is highly reduced. This provides a rough estimated of the denoised image. This estimate is then used as a reference to perform a 3D Wiener filtering on the 3D blocks from the original image, yielding the final denoised data.

The BM3D algorithm is applied individually to every band of the image, except the BSE image, which does not exhibit a significant noise level. For each X-ray band, the variance of the noise has to be provided to the algorithm as an input parameter. We estimate the noise standard deviation using the median absolute deviation (MAD), which after an appropriate scaling becomes a consistent and robust estimator of the noise standard deviation.[35] For the X-ray images that account for the presence of scarce chemical elements in the sample, strong quantization effects may appear during the acquisition process, making the noise power estimation using medians more difficult, since only a few equally spaced gray levels are present in the histogram of the image. To improve the estimation results, we add some variability to these kinds of bands by adding uniformly distributed noise with a standard deviation that is equal to half the space between two consecutive gray levels. An illustration of this step is shown in Fig. 7.

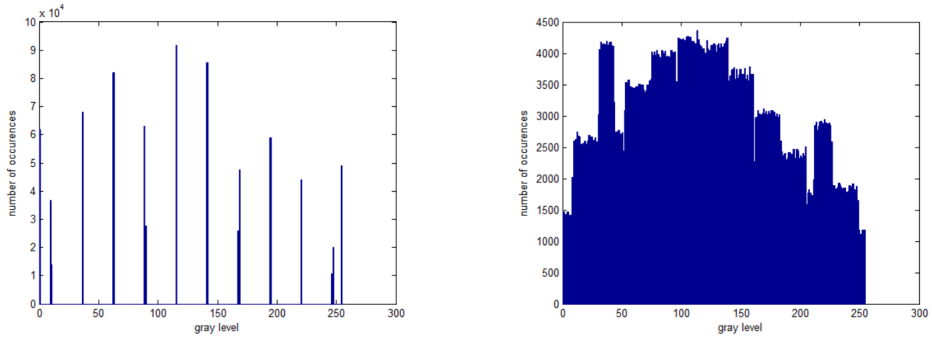

Figure 7 The histograms of a heavily quantized band, before and after addition of uniform noise for a better variance estimation with MAD.

4.2 Interactive Segmentation using BPTs

Now that the image has been denoised band by band, we describe the segmentation process in this section. We are using the BPT to provide a hierarchical representation of a multivariate image. First, we detail the construction of the BPT. Then we explain how we obtain a segmentation map from the hierarchical representation provided by the BPT, and after this how the user can interact with the tree structure to improve the segmentation map.

4.2.1 BPT construction

To build a BPT, as mentioned in section 3, we have to start from an initial partition of the image. If we choose to start from the original image, the leaves will be the pixels of the image. This approach leads to heavy computations, so instead, we chose to start from an initial segmentation, using a watershed algorithm. The gradient calculation was adapted to multi-valued images.[36] As the watershed algorithm tends to provide oversegmented partitions, it is well suited to form the leaves of the tree.

Now, in order to build the BPT, we have to measure the similarity between the regions. To this end, we have to define a region model, *i.e.* a descriptor for each region, and a merging criterion,

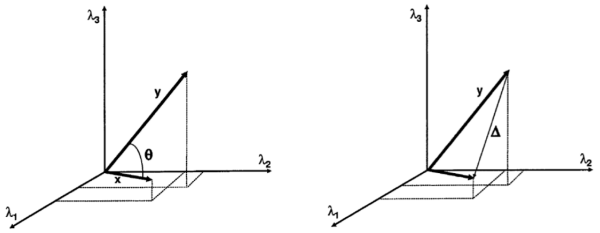

Figure 8 Two distances suited for the region model of (1), *i.e.* $\theta = SAM(x, y)$, and $\Delta = ||x - y||_2$, represented for a 3 dimensional space.[37]

i.e. a distance between region models. For multispectral images, the simplest region model $\mathbf{M}_{R_i}$ of the region R_i is given by the mean spectrum:

$$\mathbf{M}_{R_i} = \frac{1}{N} \sum_{k=1}^{N} I_i(k) \tag{1}$$

where N is the number of pixels in the region, and $I_i(k) \in \mathbb{R}^L$ with L the number of bands of the image is the (vector) value of the pixel k in the region i. $\mathbf{M}_{R_i}$ is then just the mean vector of the pixels of the region. This region model can be associated with any merging criterion based on a dissimilarity measure between vectors, such as the Spectral Angle Mapper (SAM):

$$SAM(R_i, R_j) = \arccos\left(\frac{\mathbf{M}_{R_i} \cdot \mathbf{M}_{R_j}}{||\mathbf{M}_{R_i}||_2 \times ||\mathbf{M}_{R_j}||_2}\right) \tag{2}$$

with $\cdot$ the dot product and $||\cdot||_2$ the $\mathcal{L}_2$ norm. Another possibility is the Euclidean distance in $\mathbb{R}^L$:

$$d_2(R_i, R_j) = ||\mathbf{M}_{R_i} - \mathbf{M}_{R_j}||_2 \tag{3}$$

A simple geometrical interpretation of those two distances in a 3D space is provided in Fig. 8.

The region model of (1) is a first order parametric model that will be efficient if the regions are

homogeneous. However, due to the intra-class variability in our problem (we can cite for example the shadowing effects sometimes occurring for fly ashes), this assumption is not always justified. Thus we chose to use a nonparametric statistical model based on the histograms of each band, as it was proposed by Valero et al.[22] If we denote by $\mathcal{H}_l$ the histogram of the band l of the image, the region model of the region R_i is the set of L histograms:

$$\mathbf{M}_{R_i} = \{\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_L\} \tag{4}$$

Since this model is nonparametric, it does not make any assumption on the regions properties. As we use normalized histograms, ideally, the region model can be seen as the set of approximate probability density functions (PDF) of the pixels for each band. With this in mind, we can use as a merging criterion a measure of dissimilarity between PDFs, such as the Kullback-Leibler divergence, or similar measures. However, these dissimilarity measures assume that the histograms are aligned, *i.e*. each bin of one histogram is aligned with the corresponding bin in the other one. Due to quantization and noise effects in our images, it cannot be guaranteed that our histograms will be systematically aligned. Indeed, the region model used for the BPT construction is based on the histograms of each band of the image. The bin by bin comparison of histograms can be perturbed by quantization and noise. For instance, if we are to compare the histograms of one region with itself, but with different noise realizations, some salient features of the histograms, *e.g*. local maxima can be shifted by a few bins. Thus, the histograms will appear different for a bin by bin distance, while the general shape more or less the same. This phenomenon is even more important when quantization effects occur, since a small shift can be amplified by the quantization process. To tackle this, we use a cross-bin distance between histograms, the so-called diffusion

distance.[38] This distance measures the dissimilarity of the histograms at different scales, simulating a diffusion process along the histograms (viewed as temperature fields). The scale, indexed by k below, corresponds to a different stage in the diffusion process, analogous for a point in time for diffusion processes over time. The higher k, the smoother the histograms are since they underwent k Gaussian filterings. We use this to measure the dissimilarities of our histograms at different scales, which are the successive steps of the diffusion process. The distance is the sum of the dissimilarities on all the scales.

More precisely, a being a vector in $\mathbb{R}^{N_{bins}}$ (N_{bins} is the number of bins used to compute the histograms $\mathcal{H}_l$), the diffusion process is computed by applying a Gaussian filter $\phi_{\sigma_G}(a)$ to the difference between the two histograms $d_k(a)$ at the scale k (the two histograms of R_i and R_j for the band l are denoted as $\mathcal{H}_l^{R_i}$and $\mathcal{H}_l^{R_j}$):

$$\begin{aligned} d_0(a) &= \mathcal{H}_l^{R_i}(a) - \mathcal{H}_l^{R_j}(a) \\ d_k(a) &= [d_{k-1} * \phi_{\sigma_G}]_{\downarrow 2}(a), \forall k \in [\![1, K]\!] \end{aligned} \tag{5}$$

where $\downarrow_2$ denotes a downsampling of a factor 2, used to keep the same number of points at each scale, $*$ is the convolution operator, and K is the total number of scales considered. σ_G is the standard deviation (chosen constant) of the Gaussian filter ϕ. Using all the differences at all scales, we can compute the distance D_l between two histograms for the band l as the sum of the $\mathcal{L}_1$ norms of all K scales:

$$D_l(H_{R_i}^l, H_{R_j}^l) = \sum_{k=0}^{K} \sum_{s=1}^{N_{bins}} |d_k(a)_s| \tag{6}$$

where $d_k(a)_s$ is the s^{th} component of the vector $d_k(a)$. Eventually, the diffusion distance

between our two region models (our two sets of L histograms, where L is the number of bands) is the sum of the contributions of each band :

$$D(R_i, R_j) = \sum_{l=1}^{L} D_k(H^l_{R_i}, H^l_{R_j}) \tag{7}$$

Finally, to avoid that too small regions with not really meaningful region models are merged too late in the process, we choose to grant a priority at some iterations to regions whose size is smaller than 15% of the mean size of all the regions at this point.[21]

To sum up, we build a BPT from an initial partition defined thanks to a watershed algorithm. Our region model is a set of the L histograms of a region's pixels, one per band. The merging criterion used is the diffusion distance. At each iteration, the BPT construction algorithm searches the two adjacent regions in the current partition which minimize this distance. A node corresponding to the merging of these regions is created, and the process is repeated until there is only one region left.

4.2.2 BPT pruning

Now that the BPT is built, we need to obtain a segmentation map from it. To do this, we have to prune the tree. The simplest pruning strategies either perform a cross-cut of the tree at a given height, or choose the partition obtained after a certain number of mergings (or equivalently, choose to keep the partition corresponding to a certain number of regions left to merge). However, these strategies will be suited if the tree accounts perfectly for the spatial characteristics of the image. Additionally, the choice of the parameters (height, iteration number or number of regions) is not easy. We propose here a pruning strategy to segment a multivariate image using a single parameter that accounts for the segmentation's global coarseness (and a parameter for the minimum size of

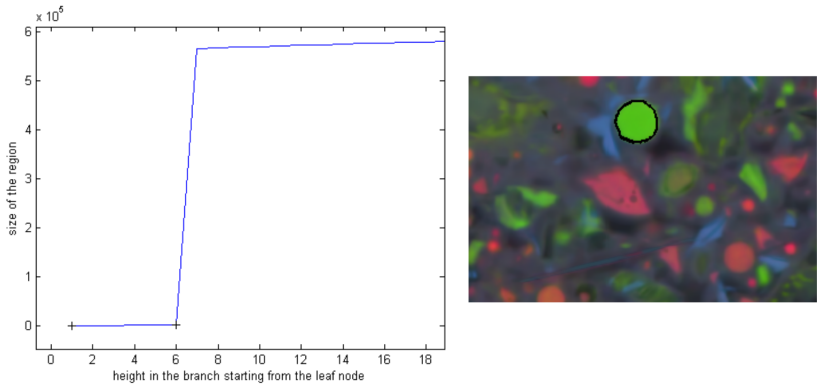

Figure 9 The rationale behind a pruning strategy based on size discontinuity in a branch: the size of the regions often present large discontinuities when they are merged with a region corresponding to a different phase. The leftmost part shows the discontinuity in the size of the regions occurring along a branch of the tree. The height is the distance from the leaf node to the node of interest. The rightmost part show the region (a fly ash) corresponding to the node situated just before the discontinuity (see the black cross on the graph), displayed on a false color composition resulting from the first three principal components.

the regions in the segmentation). It is based on the fact that meaningful properties of a node of the tree (size, mean spectrum, ...) evolve as we go up a branch[39] (the path from a leaf to the root of the tree), corresponding to more and more regions being fused to the initial leaf. An abrupt change is expected when a meaningful region merges with a very different other meaningful region (for instance two different phases that are well identified in a node of the tree that end up merging at a certain level). Then the idea will be to keep the nodes situated just before the discontinuity of the examined property, since merging the corresponding regions makes no sense. We give an example of the idea motivating this strategy in Fig. 9.

In this example, we examined the size along a branch, since it is the simplest criteria we can use. It can be adapted in many cases since even if in a given branch, a large meaningful region merges with a very different but also much smaller one, the discontinuity in size will still be present in the branch starting from the small region. However, since regions are better described by their intrinsic properties rather than only their relative sizes, there can be cases of undersegmentation

occurring. For this reason, in the proposed pruning strategy, we look for changes in the mean spectrum between the node in the branch and its sibling (the node it merges with in the tree) as we go up a branch. For simplicity, we use the SAM to measure the dissimilarity between siblings.

When there is a discontinuity in the branch (a peak in the plot above a chosen threshold on SAM), this means that two dissimilar regions (according to our threshold value) were merged. Therefore, we keep the node just before the discontinuity (before it merges with its dissimilar sibling). We also add a threshold on the the minimum size of the two regions we consider at each SAM computation to avoid that too small regions with too few pixels for their mean spectrum to be meaningful be kept in the final pruning. Obviously, the higher the threshold on the SAM, the lower the number of regions will be in the final segmentation since we will keep nodes higher in the branches because the penalty on SAM value is smaller. Besides, we will not recover regions whose sizes are smaller than the size threshold.

In the experiments, based the minimum size of the objects of interest, the threshold on size was set to 10 pixels, and based on empirical tests, the threshold on SAM was set to 0.15 rad $\simeq 8.6°$. Both thresholds were kept constant and do not need to be tuned by the user.

However, as we keep one node per branch, it is possible that the set of chosen nodes does not define a partition of the image (*i.e.* the intersection of all the selected regions is non-empty). In this case, we have to find a way to get a partition of the image. We chose to use the so-called max-tree rule strategy: we choose to keep a node in the final partition only if all the nodes in its sub-tree (the set of nodes below a given node, which define a smaller BPT) are not selected in the final partition. Otherwise, we select the two nodes below and check if the max-tree rule holds at this level. We give an example of the application of this rule in Fig. 10.

We provide an example of what happens in a branch when we apply this pruning strategy in

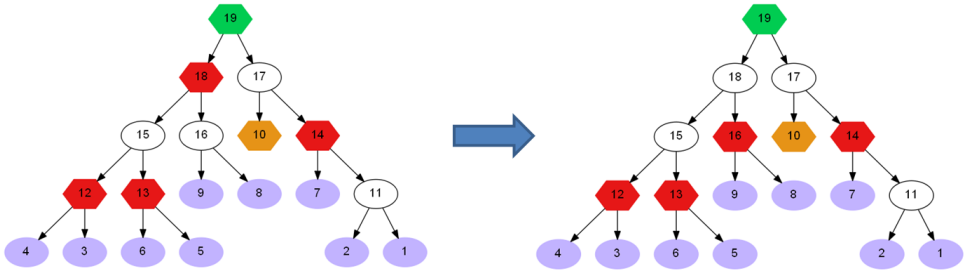

Figure 10 An example of a node selection that does not provide a partition of the image (left) and the application of the max-tree rule (right). The nodes of the tree have been colored: the root is in green, the leaves that have not been kept in the pruning are in purple, and the nodes kept in the pruning are in red (or orange if they are also leaves).

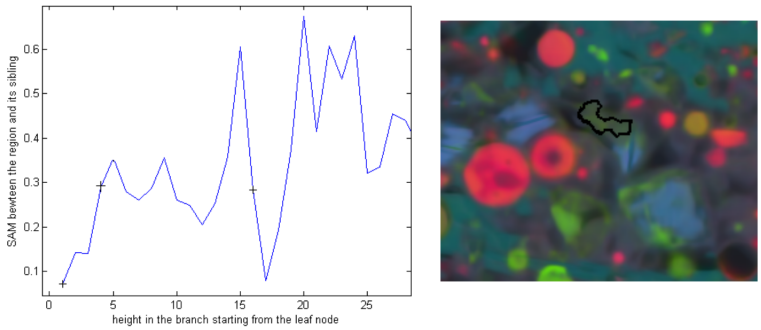

Figure 11 Evolution of SAM between nodes at a certain height in a branch and their siblings, and the region kept in the branch with the chosen pruning strategy, using a threshold on SAM of 0.15rad and a threshold on size of 10 pixels. On the graph, the black crosses indicate (from left to right) either the leaf node, the node kept by the proposed pruning strategy or the node kept considering only the size.

Fig. 11. If we used only the size criteria, we would keep nodes that are higher in this branch, because the regions at the selected level, although different, would be too similar in terms of size. The chosen strategy may be better suited in this case because we are able to discriminate better small dissimilar regions. However, contrary to the size of the node, the chosen criterion is not increasing when we go up the branch, which may somewhat complexify its interpretation.

The advantage of the chosen pruning strategy with respect to the simplest pruning strategies is that the selected nodes can lie at different levels of the hierarchy defined by the tree, depending on the branch we consider. In a way, we can say it provides a weight to the nodes of the tree, that is

to say it provides a measure of the similarity between the two regions that form the parent node. It is also easier to tune the threshold parameter on SAM and the minimum size rather than estimate the height, number of iterations or regions required to get the best segmentation possible.

4.2.3 Interactive Correction

The pruning strategy presented in the previous section allows an automatic segmentation map to be computed. However, it probably presents some imperfections to the eye of the user. Thus, we propose to integrate the user in the segmentation process by allowing him to interact with the segmentation map to correct some regions that are over or under segmented according to his visual examination. To do this, a Graphical User Interface (GUI) was implemented to allow the user to explore the tree, starting from the partition defined by the automatic pruning. The segmentation map is showed to the user next to a representation of the image (a certain band, a false color composition, ...). The user can display any region he selects and the region corresponding to the node that is just above it in the branch. Then he can choose to perform the merging and update the pruning accordingly. Similarly, he can also choose to visualize the two children of a node, and to split the region if needed. The pruning is then updated. With this simple strategy, we can use the hierarchy defined by the tree to improve the segmentation results. Plus, provided the automatic segmentation is good enough, only a few corrections (which can be often be spotted only using the BSE band) are necessary to improve significantly the segmentation map. The segmentation interaction is described in the appendix and shown in Fig. 21.

4.3 Interactive Classification using SVMs

So far, the focused was put on the computation of a good segmentation map using the BPT structure. Still, we have to keep in mind that another goal for our application is classification. In this section, we present the approach we propose to first generate an automatic classification map, an then to allow the user to interact with it to improve the classification results.

4.3.1 SVM classification

The first step of the classification process is to generate an automatic classification map using the data and the training samples we have at our disposal. The classifier we used is the Support Vector Machine (SVM), because it was successfully used in multi and hyperspectral imagery as well as for applications to cement SEM images[6,29,30] and it is able to provide good results even with a limited training set, while remaining reasonable in terms of computational burden. For the study of this paper, as we needed a multiclass classifier, we resorted to the so-called one against one approach, *i.e.* we perform binary classifications for every possible pair of classes and choose the winning class for each pixel through a voting process. Although another classifier, such as one of those reviewed in Section 3, could have been used, but we chose to use SVM for its simplicity. The objective is more to propose a framework for user interaction starting from a good enough automatic result, rather than an optimal pipeline.

The main problem of the SVM classifier is that is does not take into account the spatial structure of the images. That is why in the presence of noise, the pixelwise classification map is noisy as well. In our case, we avoid this problem since we perform the classification on the denoised data, considerably reducing the noisiness of the classification. Nevertheless, the classification map is probably not perfect. The most critical parameter in the classification is the definition of the

training samples: they have to be good representatives of the different classes and the training also needs to be balanced in order to separate the classes as well as possible. That is why the approach we propose deals with this issue by once again including the user in the process.

4.3.2 Regularization of the classification map on the segmentation map

The idea that motivates our approach is that the user is usually able to spot easily a few misclassifications in a classification map. If the user could inform the program that a certain part of the classification is wrong by correcting it, the program should be capable of learning from the corrections and adjust the classification map. That is why we need a segmentation map to identify the spatial regions of our image. We perform a regularization of the classification map on the segmentation map we obtained, that is to say, we simply assign to every region of the segmentation map the majoritary class among its pixels. This way, using a second GUI, a user will be able to select the classified regions of the segmentation map.

4.3.3 Interactive correction of the training samples

The objective of this step is to use the visual examination from the user to improve the classification map. Using a GUI, the user can select every region on the reguralization of the SVM results on the segmentation map. The user is asked to select the misclassified regions that he spots, comparing the classification maps with a representation of the data (training image, false color composition, ...), and to assign the corresponding pixels to the correct class. But instead of correcting the classification map, we choose to use the information provided by the user to improve the training samples, thus allowing the classfifier to learn from the corrections. To do so, a certain percentage of the pixels of this region are randomly selected and added to the training set with a

label that corresponds to the corrected class. The percentage has been set equal to 5% to allow the corrections to impact the classifications without unbalancing the training. It is also possible for the user to tune this parameter at every moment. By repeating this operation on a few misclassified regions, we allow the training set to be more representative of the classes and to learn from its previous mistakes. Then a new SVM classification map is generated from the new training set. If the result still remains unsatisfactory for the user, it is possible to continue the corrections until the classification gets accurate enough. It is also possible to reinitialize the corrections in case of a mistake from the user or if the training becomes unbalanced after too many corrections. The classification interaction is described in the appendix and shown in Fig. 22.

5 Validation of the method

In this section we show the results of the experiments performed on synthetic and real datasets to validate the proposed approach.

5.1 Synthetic data

5.1.1 Data generation

We generated two synthetic datasets keeping in mind that they have to be close enough to a real dataset for the results to be meaningful. To do so, in both cases (one dataset for segmentation evaluation and one for classification evaluation) we selected an output classification map of the proposed approach on a real dataset (the one of Fig. 3 in the case of classification), which we considered to be the ground truth for the synthetic data. By assigning slow varying gray level values to each region in one band, we generated a synthetic image with realistic geometrical structures. To simulate the coarser spatial resolution of the X-ray bands, we applied a blurring gaussian filter

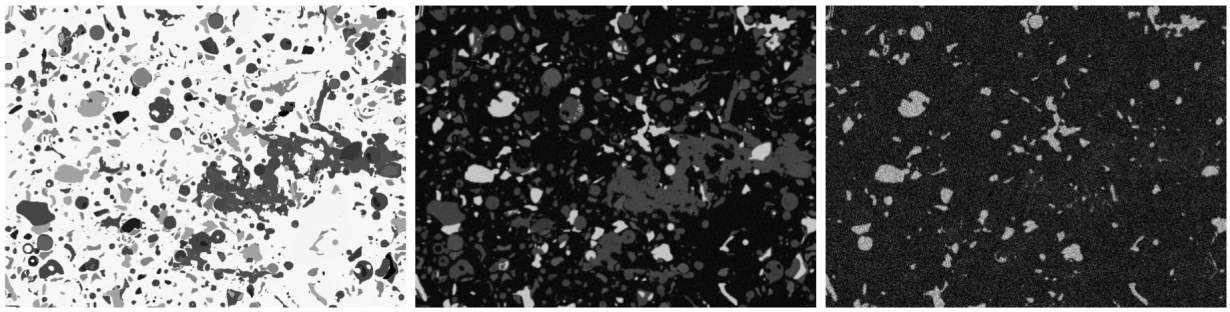

Figure 12 Some bands of the synthetic dataset used for classification evaluation. A synthetic band mimicking the BSE band (left), a normal X-ray band (middle) and a quantized X-ray band (right) are shown.

to these bands. We added a 10dB gaussian noise to some bands, and a 0dB noise to some others. Indeed, we wanted to reproduce the quantization effects present in lower SNR bands. So, to simulate the quantization effects present in certain bands, we also uniformly quantized the low SNR X-ray bands so that they only comprise around 10 gray levels. We show the generated "BSE" image and two "X-ray" bands in Fig. 12. In the classification results, the names of the real classes are kept although their response in the synthetic X-ray bands do not always match the true ones (but their shape and geometry do: *e.g.* Fly Ashes keep their round or crescent shape but are associated with hypothetical chemical elements here). For the classification dataset, the training pixels used are at the same locations as in the image of Fig. 1.

5.1.2 Segmentation evaluation

In this section, we compare the results of the automatic segmentation result provided by the BPT pruning to the output of the segmentation from the mean shift algorithm.[15] This is a flexible algorithm which was successfully used in different contexts. The implementation we used could only handle color images, so in order to use most of the information provided by the data, we ran this algorithm on the first three principal components of the image. Besides, for a fair comparison, both algorithms were run on the denoised data. We compared the outputs of the segmentation results

to a ground truth obtained by converting the ground truth classification map into a segmentation map, by labelling the connected components of this image. For the quantitative results, we used three metrics, whose properties and implementations are discussed in the paper by Cardoso and Corte-Real.[25] The symmetric distance $d_{sym}(R, P)$ between a candidate partition P and a reference partition R, and the assymetric distances $d_{asym}(R, P)$ and $d_{asym}(P, R)$ between a partition and a reference. The symmetric distance can be defined as the number of pixels to be removed to the support of the image so that the two partitions are identical (up to a permutation of the labels), possibly normalized by its maximum possible value for two partitions ($N - 1$, where N is the number of pixels in the image), so that the measure values are between 0 and 1. This measure defines a distance between partitions which we can use for comparing the outputs of different segmentation algorithms when a reference is available, as in this case. The assymetric distance $d_{asym}(R, P)$ (here the order of the arguments matters) is defined as the number of pixels to remove to the support of the image so that the partition P is a refinement of partition R, meaning that every region of P is contained in a region of R. This distance enables us to compare two partitions, but without being sensitive to oversegmentation. This means that the distance between a partition and a refinement of it will always be zero. This measure is useful to our application if we keep in mind that region corrections for the classification will be much more cumbersome in case of undersegmentation than oversegmentation. We also provide the value of $d_{asym}(P, R)$, which defines a new measure insensitive to undersegmentation.

The visual results of the segmentation algorithms are provided in Fig.13, and the quantitative results are gathered in Table. 1.

From the visual results, we can see that both tested algorithms recover more or less the shape of the regions in the ground truth image. However, exact recovery is made difficult in both cases because

Measure (normalized) / Algorithm	$d_{sym}(R,P)$	$d_{asym}(R,P)$	$d_{asym}(P,R)$
Mean Shift	0.5931	**0.0899**	0.5802
Proposed	**0.3294**	0.1232	**0.2980**

Table 1 Quantitative results for the segmentation on a synthetic dataset.

of the coarser resolution of the images mimicking X-ray bands. It seems that overall the proposed approach tends to undersegment the image slightly more than the mean shift output, though this is not always true locally. Conversely, the mean shift output seems overall more oversegmented, which is not critical in our application since the segmentation is a tool for the region correction in the interactive classification process.

The quantitative results confirm these observations. The symmetric distance with the reference is lower in the case of the proposed method (with the thresholds set to a fixed value so as not to be tuned between datasets), meaning that the segmentation is overall closer to the reference. However, the assymmetric distance $d_{asym}(R,P)$ (insensitive to oversegmentation) is slightly in favor of the mean shift algorithm, meaning that without considering oversegmentation, the mean shift segmentation is closer to the reference. On the contrary, finally, the distance $d_{asym}(P,R)$, insensitive to undersegmentation, is better for the proposed approach. Note that even if oversegmentation is not as critical as undersegmentation in our application, the proposed approach still has the advantage of being a hierarchial segmentation. This means that user corrections is possible and easy in order to improve the segmentation in undersegmented regions, and also that the scale of the regions in the results of the proposed approach can still be refined depending on the value of the two thresholds for the pruning. These results show that the proposed approach is competitive with other segmentation algorithms and that it is a satisfactory automatic result which is an instance of a hierarchical segmentation process.

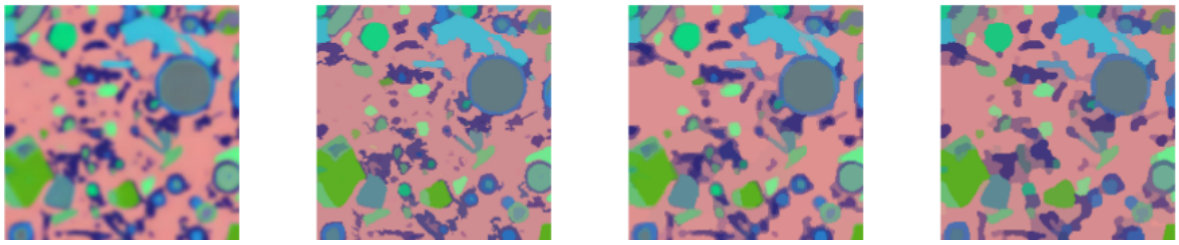

Figure 13 Segmentation results on a synthetic dataset. From left to right: Color composition (PCA), Ground truth segmentation, Results of the Mean Shift algorithm, and results of the proposed BPT pruning strategy.

Method \ Phase	Alite	Belite	Portlandite	Hydrates	Sulfoaluminates	Inners	Quartz	Fly Ashes	Porosity	Overall
SVM-MRF	**96.94**	92.16	**98.17**	**99.70**	91.86	80.74	95.42	72.20	84.62	90.39
Proposed	90.33	**92.37**	95.63	99.51	**94.04**	**85.57**	**97.44**	**93.31**	**90.58**	**94.19**

Table 2 Quantitative results on the synthetic dataset.

5.1.3 Classification results

We ran the SVM-MRF algorithm and the proposed approach on the synthetic data. For the proposed approach, 10 region corrections were performed. The visual results (Fig. 14) show that the proposed approach misses less small structures, such as for Fly Ashes (whose phase percentages are known to be underestimated by SVM-MRF due to their sometimes small size), and porosity. The quantitative results, given in Table 2, show that the proposed approach significantly outperforms the SVM-MRF method, overall and on specific components. The main improvement comes on the fly ashes, with a improvement of more than 20%. The only loss in some phases occur for scarce or relatively scarce phases (Alite and Portlandite), or a slight loss on the hydrates, which form the background of the images. Since corrections are performed on some phases, it is normal that this adds a very small amount misclassifications in the background. All others classes are better, or much better classified by the proposed approach, which tends to confirm the relevance of the classification framework we propose.

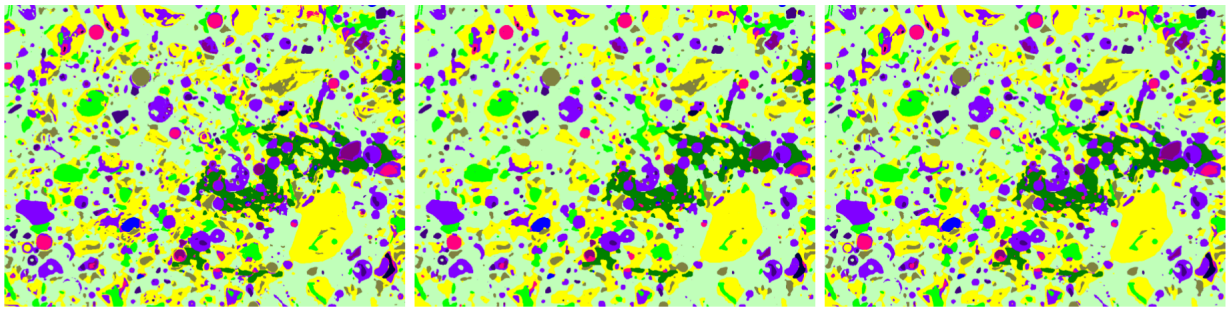

Figure 14 Classification results on a synthetic dataset. The true classification map is shown on the left, while the results of SVM-MRF and the proposed approach are shown on the middle and on the right, respectively.

5.2 Real data

Here we describe the results of the proposed approach on a real dataset. First, we present the results of the denoising step using the BM3D algorithm. Then we show the results of the automatic segmentation using the BPT and the pruning strategy we developed for our application. Finally we present the results of the classification using the SVM and the training correction GUI, and we compare it both visually and quantitatively to the SVM-MRF method,[6] with 3×3 neigborhoods.

5.3 Denoising Results

Here we present the results of the BM3D denoising. The results on two different X-ray bands of our data are shown in Fig. 15. We compare the denoising results with the ones obtained using a 3×3 median filter. We can see that BM3D allows to reduce the noise much more importantly than the median filter, without significantly altering the spatial structure of the image. Obtaining this level of smoothness from a large median filter would be on the contrary highly detrimental to the edges of the images. The results are particularly good visually on bands presenting quantization effects such as the Cl band.

We can also visualize the results of the denoising on the training set: in Fig. 16 we show a scatter plot of the training samples corresponding to Fig. 3, in a 3D space corresponding to

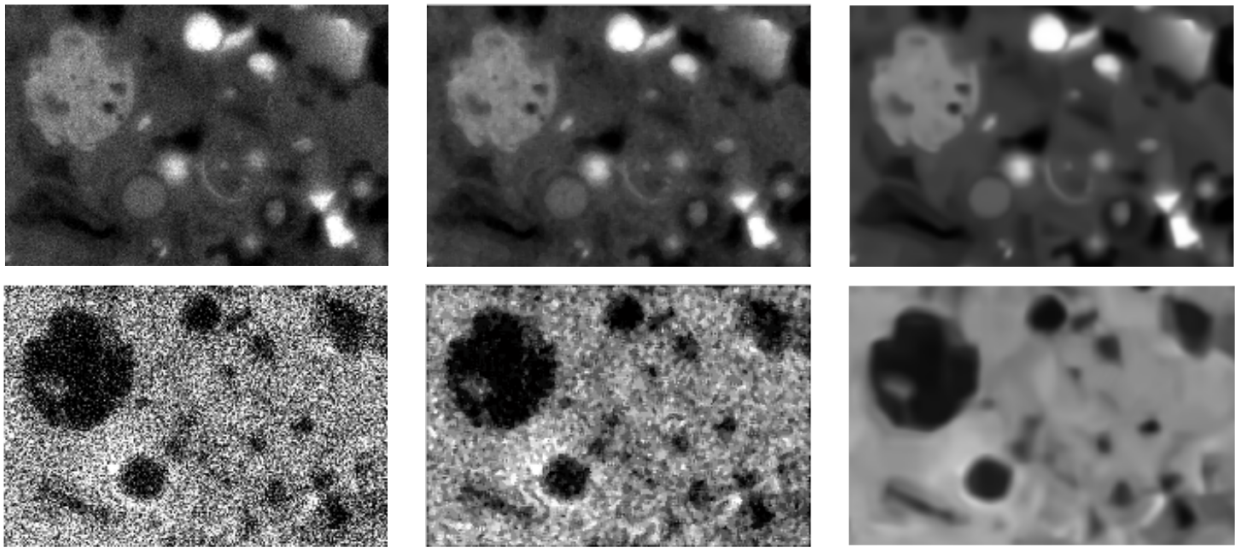

Figure 15 Results of the denoising on two bands (leftmost column) using a 3×3 median filter (middle column) and the BM3D algorithm (rightmost column) on small parts of two X-ray bands: Aluminium (top row) and Chlore (bottom row) of a cement sample.

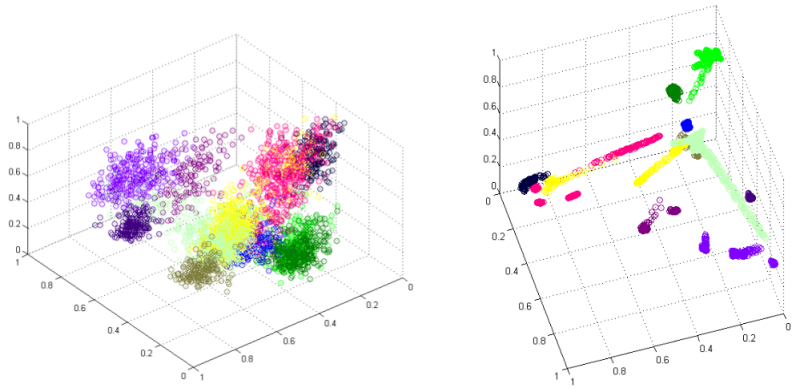

Figure 16 Scatter plots of the training set of Fig. 3 on the first 3 principal components before (left) and after (right) the denoising steps.

the first 3 principal components of the set. We can see that the denoising allows a much better separation of the classes than with the raw data, with which many classes overlap. However, the fact that the training samples present less variability after the denoising may cause the training set to become less informative in itself. In case this happens, the problem will be solved anyway during the training correction step, during which it will be possible to add new training instances.

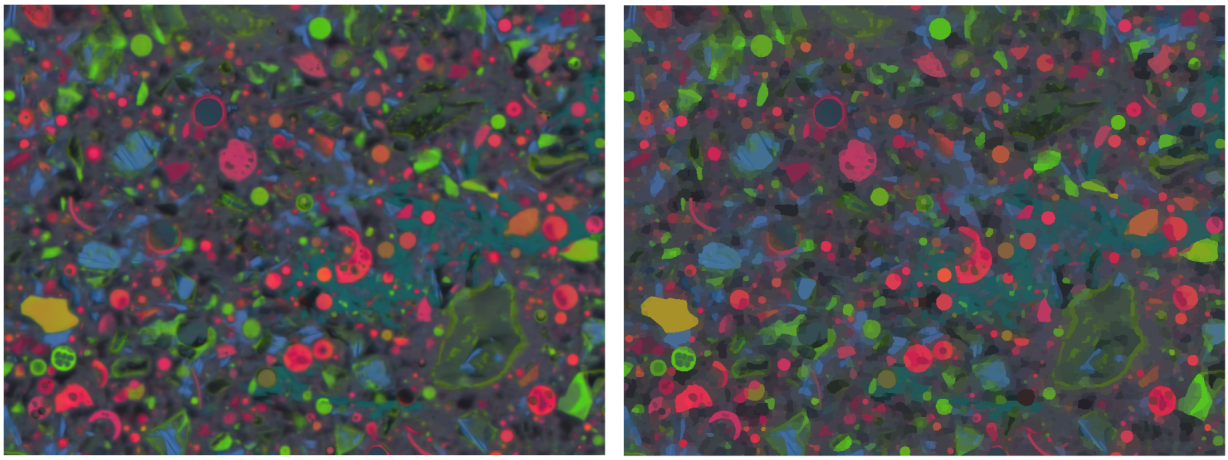

Figure 17 Segmentation results on the real dataset. A false color composition (PCA) is shown on the left. The automatic segmentation map obtained by the proposed approach is shown on the right (Every region is represented by its mean color).

5.4 Segmentation Results

Now we present the results of the automatic segmentation described in section 4.2.2. The visual results can be seen in Fig. 17. We represented as a reference the false color composition corresponding to the first three principal components of a PCA performed on the denoised multispectral image. We can see that the vast majority of salient spatial elements are correctly recovered by our segmentation map. Even the smallest circular regions, corresponding to micronized fly ashes are present in the BPT structure and in most cases correctly detected by the proposed pruning strategy. Only some of the smallest regions are merged with small surrounding regions, probably because they were not separated in the first place by the watershed algorithm. The latter is still extremely useful to limit the computation time. Besides, it has to be kept in mind that the segmentation map can be corrected and improved by the user, provided the BPT's structure allows the desired merging or splitting.

5.5 Classification Results

The results of the classification process after the interaction with the user are presented in this section. First we show the classification maps, in comparison with those obtained by a basic SVM classification with no denoising and the SVM-MRF method. Then we will present quantitative results based on the comparison of the phase percentages of 1 day-old cement paste computed using the classification maps and the reference percentages coming from the knowledge of the cement composition.

5.5.1 Visual results

We present in Fig. 18 the BSE image used for the experiment, with the training pixels displayed in color (see Fig. 3 for the information on the classes), as well as the results of the segmentation coming from SVM (without the denoising step), SVM-MRF and the proposed approach. Note that contrary to the images tested for the quantitative results part, this image depicts a 91-old cement paste, in which the different components have started the hydration reactions.

We can see that the classification results obtained with the proposed approach respect the spatial structure of the original image, seen in the BSE band. The classification method we propose allows to beneficiate from the conservation of the spatial features provided by the SVM without the drawback of getting noisy classification maps, as the pixelwise classifier obtains in for noisy data. It also allows us to reach the smoothness shown in the SVM-MRF result, but without losing the information on the edges, and it is able to recover even the smallest particles in the classification map. The reasons for this are twofold: the denoising step smoothes the result provided by the SVM and the training correction using the segmentation map allows to include the smallest particles and the originally misclassified regions in the training set. In particular, a close look to the images

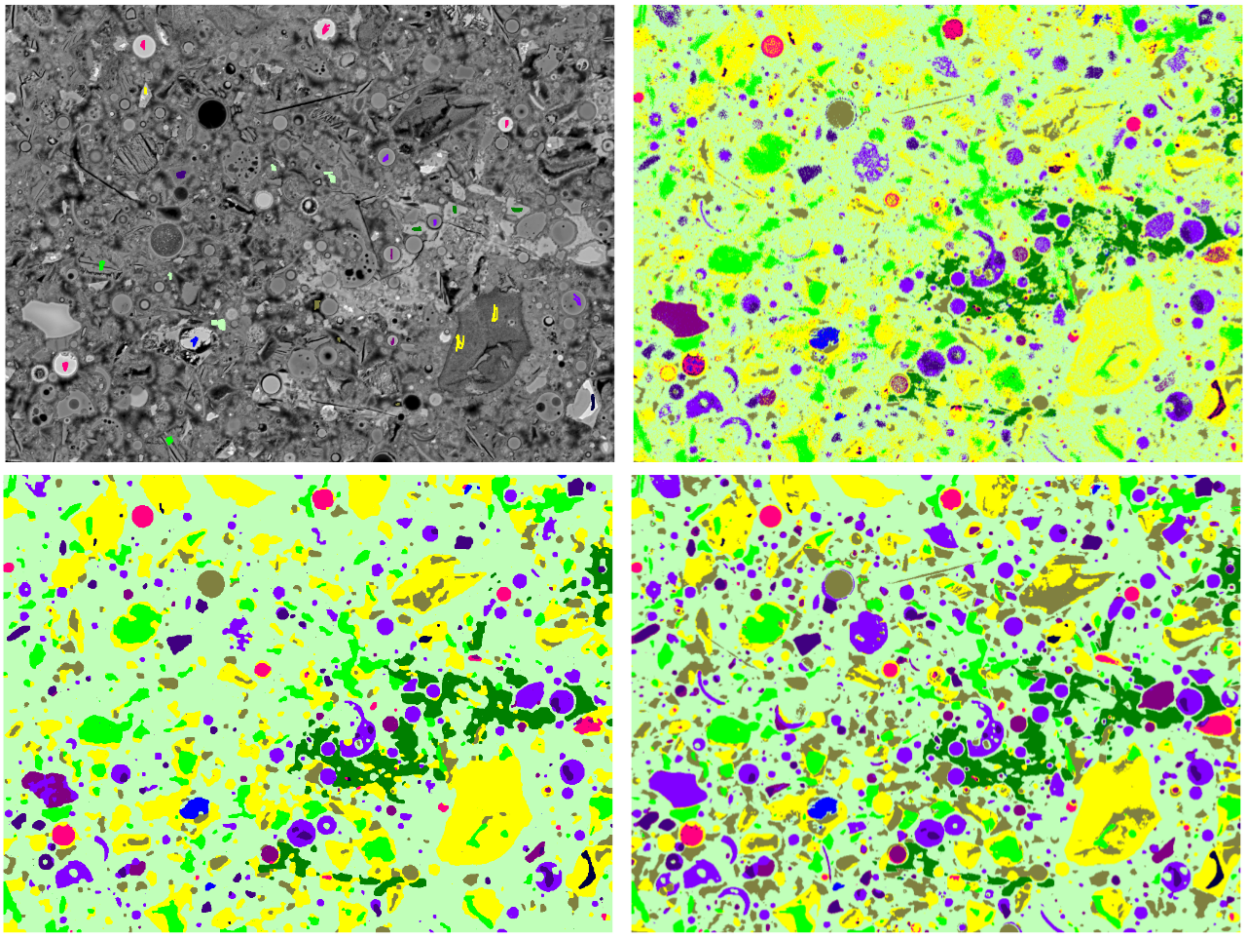

Figure 18 Classification maps of the proposed method in comparison with SVM and SVM-MRF. For the information on the classes, see Fig. 3. On the top row are displayed the training image and the SVM classifation map on the raw data (without denoising). On the bottom row we show the results of SVM-MRF (left) and of the proposed approach (right), after the integration of a dozen misclassified regions in the training set.

shows that we are able to detect much more small micronized fly ashes than with the SVM-MRF or even the SVM method.

5.5.2 Quantitative Results

In this part, we present a way to quantitatively assess the results of the classification method we propose. It is based on the processing of cement samples one day after the solidification of the paste. At this point, we can consider that some of the additives (fly ashes, limestone fillers, pozzolana and slag for instance) have not begun to react in the paste. Knowing the percentage accounted for each of these components in several cement pastes, it becomes possible to have a reference for the comparison with the experimental phase percentages, *i.e.* the percentage of the image pixels for each class. A complete SEM acquisition from a cement sample represents at least 30 tiles of the same size as the images we have presented throughout this paper ($1024 \times 768 \times 15$). The training and the interactive part are only performed on one of them, so that the corrected training can be used to perform automatic SVM classifications on all tiles, after the denoising step. Then we have to average the percentages found for each tile. To make sure the obtained percentages are statistically meaningful, we plotted in Fig. 19 the relative variation of the accumulated means of the phase percentages every time a new tile is processed, for a sample comprising around 40 image tiles. This means a point in the plot for a given class corresponds to the relative variation between the mean phase percentages of this class when n and $n + 1$ tiles have been processed. We can see in this plot that for each class, the relative variation quickly decreases to small values when the processed part of the complete sample becomes larger. After 25 images, the variations are almost never larger than a few points. From this, it seems that having at least 30 tiles at our disposal is enough to get statistically meaningful results.

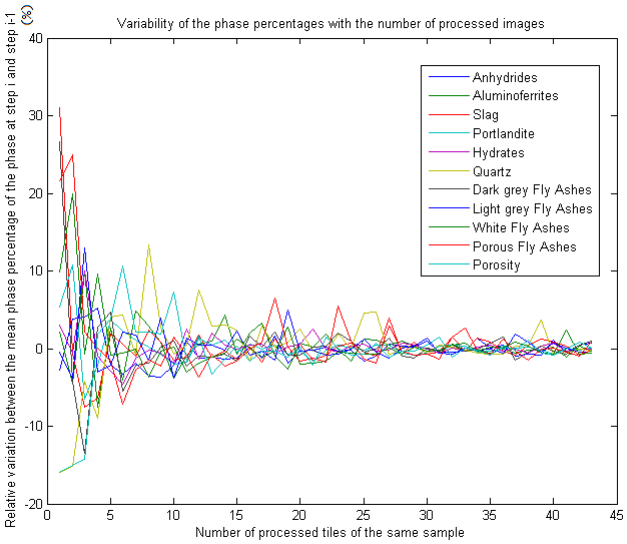

Figure 19 Relative variation between the accumulated mean phase percentages of each class of the image of Fig. 3 throughout the processing of a full sample.

The comparison of the average phase percentages of each phase of interest has been performed for several cement systems. We computed the phase percentages of the classes corresponding to meaningful groups (for instance, there can be several types of fly ashes defined in the training set that we group in one single class) whose quantity put a priori in the cement is known. These groups can be, depending on the samples: fly ashes, slag, limestone filler, or several of those if they are present together in the sample. These compounds are not supposed to have reacted after one day. Finally, we can plot the experimental values against the theoretical ones, where each symbol in the plot accounts for the results concerning one of the aforementioned groups, after the processing of one full sample (from 25 to 50 images). This has been done in Fig. 20 , with the results from the proposed method (in black) and the SVM-MRF method (in color). We also plotted the first bisector as a reference. We also show the same results in Table 3.

From the plot, we can see that our method performs better than the SVM-MRF method pro-

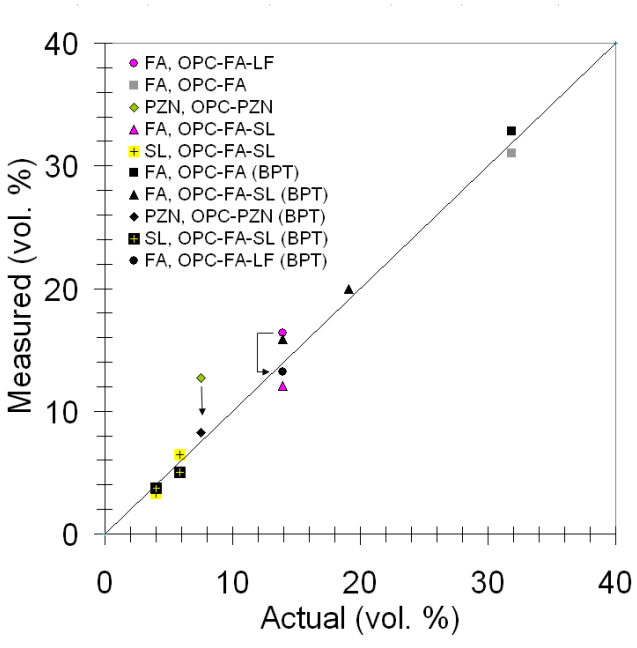

Figure 20 Experimental phase percentage values of cement components 1 day after solidification plotted against theoretical values for several cement systems (represented by different symbols). The optimal values lie on the first bisector. FA: Fly Ashes, SL: Slag, PZN: Puzzolana, LM: Limestone Filler, OPC: Ordinary Portland Cement. The black symbols correspond to the proposed approach (BPT) and the colored symbols to SVM-MRF.

vided the corrections of the training are reasonably carried out. We can see that the BPT approach performs better than SVM-MRF in 3 cases out of 7, comparably in 3 cases out of 7, and worse in only one case. The main improvement can be seen on two cases: FA in a ternary OPC-FA-LF mixture, and PZN in an OPC-PZN mixture (which corresponds to the arrows in Fig 7). The small size of the FA and the complexity of the mixture in the first case and the presence of clay in the second are better handled by the proposed approach. Micronized fly ashes were better identified

Method \ Phase	FA in OPC-FA-LF	SL in OPC-FA-SL	SL in OPC-FA-SL	FA in OPC-FA-SL	FA in OPC-FA-SL	PZN in OPC-PZN	FA in OPC-FA
BPT	**13.22**	**3.70**	5.00	15.88	19.97	**8.25**	32.84
SVM-MRF	16.40	3.27	**6.45**	**12.05**	**19.95**	12.70	**31.00**
Theory	13.90	3.97	5.84	13.92	19.10	7.49	31.87
Absolute difference with theory (BPT)	***0.68***	***0.27***	0.84	1.96	0.87	***0.76***	0.97
Absolute difference with theory (SVM-MRF)	2.50	0.70	*0.61*	1.87	0.85	5.21	0.87

Table 3 Phase percentages obtained by the proposed approach (BPT) confronted to those of SVM-MRF. OPC: Ordinary Portland Cement, FA: Fly Ashes, PZ: Pozzolan, SL: Slag, LF: Limestone Fillers. In the first two rows, bold values indicate the best result. In the last two rows, bold and italic values indicate better performance for the proposed approach, the italic value correspond to a better performance for SVM-MRF, and in the columns without bold or italic values, both algorithms perform similarly (absolute difference between the two differing by 0.1 or less).

by the proposed method, especially in the ternary OPC-FA-LF mixture, which confirms the visual results. This suggests the potential of the interactive BPT approach as a robust technique for the classification of complex mixtures.

Finally, we give some information regarding the computational burden of both methods. For an image similar to the ones showed in this paper, and whose training has to be improved interactively, computing the segmentation and classification maps (without considering the few minutes necessary for corrections) with the proposed approach takes roughly the same time (around 10 min) than computing a classification map with the SVM-MRF method. However, the proposed method is computationally more efficient when automatically classifying a batch of images of the same cement sample, once the training has been corrected. Indeed, for such images, only classification maps have to be computed, which require 2 or 3 min per image, while 10 are necesserary for SVM-MRF. The computation time for segmentation is indicative since the it also depends on the spatial complexity of the images.

6 Conclusion

In this paper, we presented a new method for the segmentation and classification of multispectral SEM images of cementitious materials using an interactive process. The proposed method is a pipeline involving severa image processing tools in order to provide automatic segmentation and classification maps, as well as defining a framework in which the user can easily improve the results with a few interactions with the program. We were first able to denoise efficiently the data. For the segmentation part, we used a hierarchical region based representation of the image using Binary Partition Trees. We developed a pruning strategy to recover an automatic segmentation from the BPT. Then we developed a process in a GUI with which the user can explore the tree starting from

the automatic segmentation to improve the results. The automatic segmentations recovered by the algorithm seem to account correctly for the spatial structure of the data, based on visual examination. The classification part is carried out using Support Vector Machines. Once again, once a supervised classification map has been computed using some training sample define by an expert, the user can use a regularization of the classification map on the segmentation result to integrate initially misclassified pixels to the training set, thus improving its robustness and representativity of the classes. The classification maps obtained after the interactive part of the process are practically noiseless and seem to preserve the spatial information of the data as well as detecting even the smallest particles, which is a significant improvement with respect to the SVM-MRF method whose results were compared with ours. On a more quantitative point of view, we were able to compare the accuracy of the segmentation and classification maps we obtained on synthetic data, and on real data for unreacted components of some cement samples. Once again, the proposed approach gives results that are close to the theoretical values and outfperorms the SVM-MRF method. The relevance of the method we propose makes it a good and flexible solution to estimate phase percentages of the same cement sample over time, and consequently compute hydration degrees of the components.

Even though the proposed method seem to be relevant to tackle the cementitious material SEM images segmentation and classification problem, we can think of future developments for this work. First, we saw that the training samples are the most sensitive parameter of every supervised classification technique, so it should be interesting to study how we should select the training samples, *i.e.* how to choose the ones who are the most informative and allow us to separate the data in an optimal way. This process, called Active Learning, is currently receiving a lot of attention in the domain of Machine Learning, and could be an alternative to the interactivity with the user in

the proposed approach. Finally, we could explore the potential of the hierarchical representation provided by BPTs to try to get better segmentation maps: the idea would be to interact with the user in a first step and then try to infer rules from his decisions and propagating them to other parts of the tree. Then we could think of developing new pruning strategies for segmentation that match better the visual interpretation. New pruning strategies specifically dedicated to classification could also be investigated as another line of attack or to be combined with the interactive classification process proposed in this paper.

Appendix

In this appendix are presented the two GUIs used for the improvement of the correction and classification parts of the process (Figs. 21 and 22). The image of Fig. 21 shows the segmentation interface. On the left can be shown several images of interest: the different bands, a false color composition of the first three principal components, the initial segmentation map... When the user left clicks a region on the right, the corresponding region is displayed, as well as the sibling node. The user can merge the two regions or leave them as is. With a right click, the two children regions of a node are shown, and the user can split the region into its two children. In the case of Fig. 21, the user left clicked a round region (fly ash) and was proposed to merge it with the large region in blue, which is not a good choice in this case.

The classification interface is shown in the image of Fig. 22. On the left is displayed an image of interest (band, color composition, original SVM classification...). On the right is shown the regularization of the SVM classification on the segmentation map. The user can click any region on the left image, or any part of the left image to display the corresponding region. Then using the boxes at the bottom left corner, he can choose the label corresponding to the right class, which

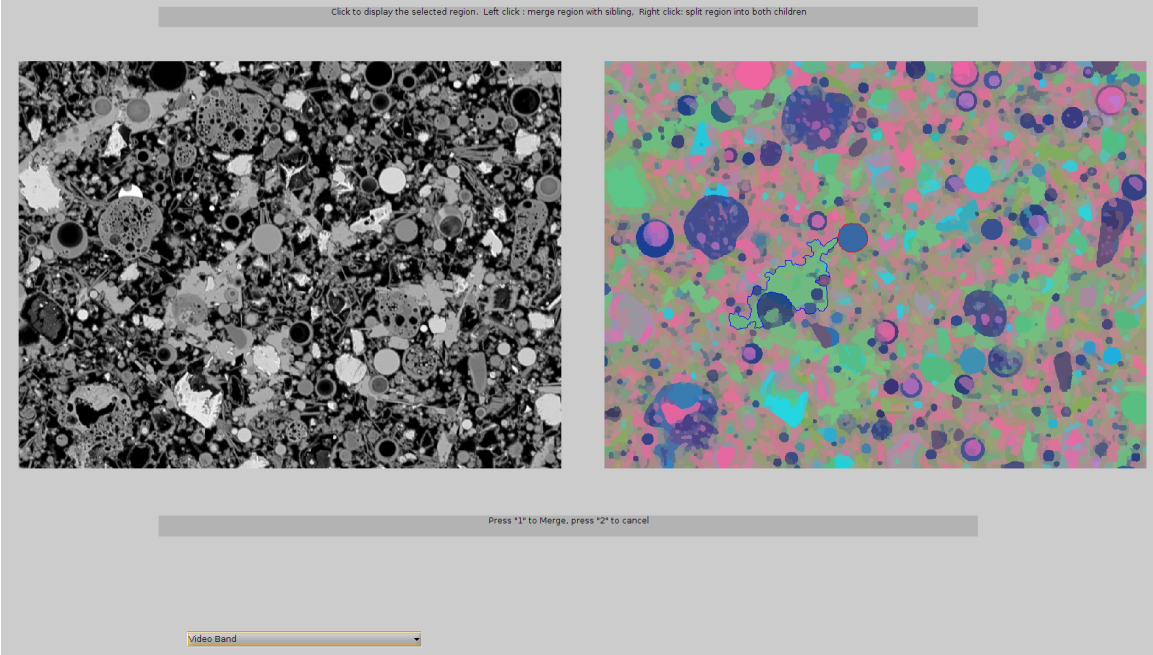

Figure 21 The segmentation GUI.

will change its color on the right image and add a certain percentage of the pixels of this region to the training set (5 by default). When the user has finished his corrections, he can generate a new classification map. He can then (if more corrections are needed) update the interface with new results (and generate a new regularized classification) in order to continue the training improvement. Here the clicked region should be changed to class 11 (White fly ashes).

References

1 Q. Wu, F. A. Merchant, and K. R. Castleman, *Microscope Image Processing*, Academic Press (2008).

2 J. Goldstein, D. Newbury, D. Joy, C. Lyman, P. Echlin, E. Lifshin, L. Sawyer, and J. Michael, *Scanning Electron Microscopy and X-Ray Microanalysis (3 ed)*, Springer (2003).

3 M. B. Haha, K. de Weerdt, and B. Lothenbach, "Quantification of the degree of reaction of fly ash," *Cement and Concrete Research* **40**, 1620–1629 (2010).

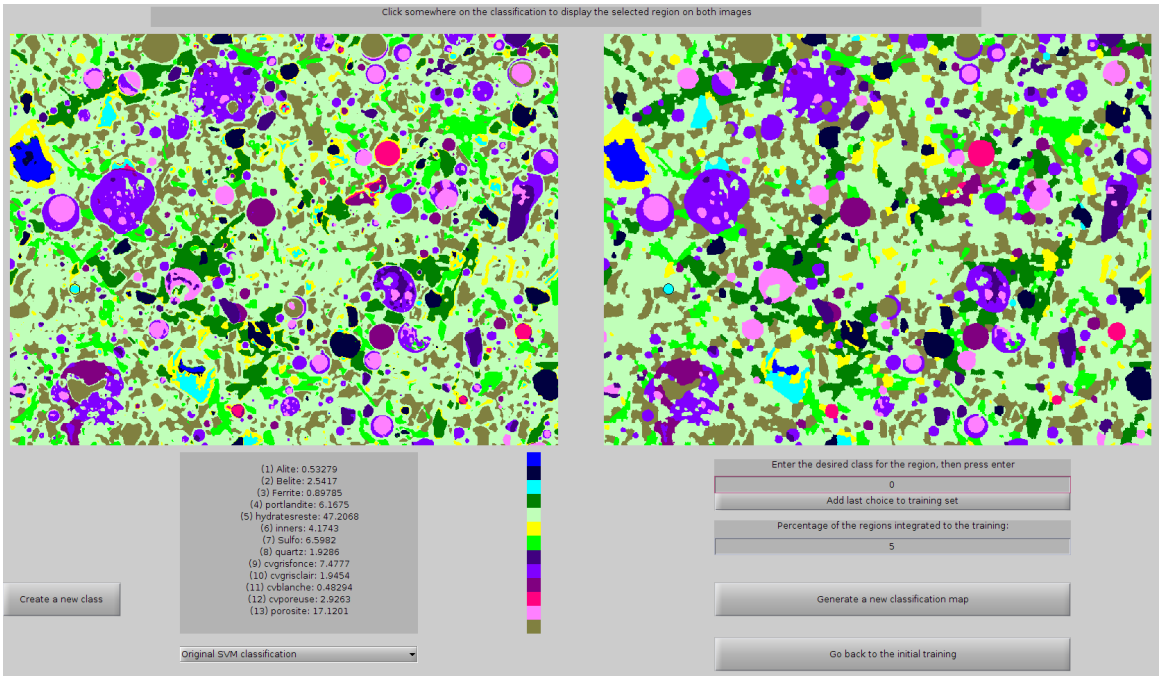

Figure 22 The classification GUI.

4 L. Drumetz, M. Dalla Mura, S. Meulenyzer, S. Lombard, and J. Chanussot, "Semi-automatic classification of cementitious materials using Scanning Electron Microscope images," in *Proc. SPIE 9534, Twelfth International Conference on Quality Control by Artificial Vision 2015, 953403*, **9534,**, (Le Creusot, France) (2015).

5 J. R. C. P. E. Stutzman, "Specimen preparation for scanning electron microscopy," in *Proceedings from the Twenty-First International Conference on Cement Microscopy*, 25–29, National Institute on Standards and Technology, (Las Vegas, Nevada, USA) (1999).

6 S. P. Meulenyzer, J. Chanussot, S. Crombez, and J. J. Chen, "Spectral spatial image processing strategies for classifying multipsectral sem-eds x-ray maps of supplementary cementitious materials," in *14th Euroseminar on Microscopy Applied to Building Materials*, (Helsingør, Denmark) (2013).

7 D. P. Bentz and P. E. Stutzman, "Sem analysis and computer modeling of hydration of port-

land cement particles," in *Symposium on Petrography of Cementitious Materials*, 60–73, (Atlanta, GA, USA) (1993).

8 P. E. Stutzman, "Scanning electron microscopy imaging of hydraulic cement microstructure," *Cement and Concrete Composites* **26**, 957–966 (2004).

9 A. G. Silva, R. de Alencar Lotufo, and F. C. Flores, "Classification of microstructures by morphological analysis and estimation of the hydration degree of cement paste in concrete," in *Proceedings of the XV Brazilian Symposium on Computer Graphics and Image Processing*, (2002).

10 S. Beucher and F. Meyer, *Mathematical Morphology in Image Processing*, Marcel Dekker (1992).

11 C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning* **20**, 273–297 (1995).

12 A. Blake, P. Kohli, and C. Rother, *Markov Random Fields for Vision and Image Processing*, MIT Press (2011).

13 R. C. Gonzalez and R. E. Woods, *Digital Image Processing, third edition*, Prentice Hall (2007).

14 D. Comaniciu and P. Meer, "Mean shift: A robust approach towards feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, 603–619 (2002).

15 D. Martins, V. G. Josa, G. Castellano, and J. B. da Costa, "Phase classification by mean shift clustering of multispectral material images," *Microscopy and Microanalysis* **19**, 1266–1275 (2013).

16 N. M. Nasab, M. Analoui, and E. J. Delp, "Robust and efficient image segmentation ap-

proaches using markov random field models," *Journal of Electronic Imaging* **12**(1), 50–58 (2003).

17 P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based segmentation," *International Journal of Computer Vision* **59**(2), 167–181 (2004).

18 Y. Zheng, B. Jeon, D. Xu, Q. Wu, and H. Zhang, "Image segmentation by generalized hierarchical fuzzy c-means algorithm," *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology* **28**(2), 961–973 (2015).

19 P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation and information retrieval," *IEEE Transactions on Image Processing* **9**, 561–576 (2000).

20 S. Valero, P. Salembier, and J. Chanussot, "Comparison of merging orders and pruning strategies for binary partition tree in hyperspectral data," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2565–2568 (2010).

21 S. Valero, P. Salembier, and J. Chanussot, "Hyperspectral image representation and processing with binary partition trees," *IEEE Transactions on Image Processing* **22**(4), 1430–1443 (2013).

22 S. Valero, *Arbre de partition binaire : Un nouvel outil pour la représentation hiérarchique et l'analyse des images hyperspectrales*. PhD thesis, Université de Grenoble, Universitat Politècnica de Catalunya (2011).

23 F. Flores and R. de Alencar Lotufo, "A new hierarchical decomposition applied to object segmentation in image sequences: The uniform decomposition," in *Graphics, Patterns and Images (SIBGRAPI), 2010 23rd SIBGRAPI Conference on*, 156–163 (2010).

24 H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *computer vision and image understanding* **110**(2), 260–280 (2008).

25 J. S. Cardoso and L. Corte-Real, "Toward a generic evaluation of image segmentation," *Image Processing, IEEE Transactions on* **14**(11), 1773–1782 (2005).

26 Q. Ding and M. Colpan, "Decision tree induction on hyper-spectral cement images," *International Journal of Information and Mathematical Science* **2**(3), 169–175 (2006).

27 C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer (2007).

28 C. N. Silla Jr and A. A. Freitas, "A survey of hierarchical classification across different application domains," *Data Mining and Knowledge Discovery* **22**(1-2), 31–72 (2011).

29 F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing* **42**, 1778–1790 (2004).

30 Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters* **7**, 1430–1443 (2010).

31 R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **34**, 2274–2282 (2012).

32 M. Qiu, Z. Li, H. Zhang, C. Xie, and J. Zhang, "Joint sparsity matrix learning for multi-class classification applied to face recognition," *Journal of Electronic Imaging* **23**(3), 033007 (2014).

33 Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *Geoscience and Remote Sensing, IEEE Transactions on* **49**, 3973–3985 (2011).

34 K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Transactions on Image Processing* **16** (2007).

35 D. Hoaglin, F. Mosteller, and J. Tukey, *Understanding robust and exploratory data analysis*, vol. 3, Wiley New York (1983).

36 Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognition* **43**, 2367–2379 (2010).

37 N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Transactions on Geoscience and Remote Sensing* **42**, 1552,1565 (2004).

38 H. Ling and K. Okada, "Diffusion distance for histogram comparison," *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference* **1**, 246–253 (2006).

39 G. Tochon, J. Féret, R. Martin, R. Tupayachia, J. Chanussot, and G. Asner, "Binary partition tree as a hyperspectral segmentation tool for tropical rainforests," in *Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE International, (Munich, Germany) (2012).

Lucas Drumetz received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 2013. He is currently working towards the Ph.D. degree in signal and image processing at Joseph Fourier University, in GIPSA-lab, Grenoble,

France. His research interests include multispectral and hyperspectral image processing, as well as convex optimization techniques and their applications to remote sensing.

Mauro Dalla Mura received the M.E. degree in Telecommunication Engineering from the University of Trento, Italy, in 2007. He obtained in 2011 a joint Ph.D. degree in Information and Communication Technologies (Telecommunications Area) from the University of Trento, Italy and in Electrical and Computer Engineering from the University of Iceland, Iceland. He is currently an Assistant Professor at Grenoble Institute of Technology, France. He is conducting his research at the GIPSA-Lab.

Samuel Meulenyzer received his MSc degree in material science from Valenciennes institute of technology, Valenciennes, France, in 2001. Currently he is a senior research scientist and head of the microscopy and image analysis team at the central research center of LafargeHolcim group, Lyon, France. His research interests include microscopy and imaging of building materials, especially by using multispectral and hyperspectral image processing applied on binders and Supplementary Cementitious Materials (SCms) formulation and reactivity.

Sébastien Lombard is a research technician in the microscopy and image analysis team at the central research center of LafargeHolcim group, Lyon, France. He received his BTEC Higher National Diploma in physics in 2001 in Grenoble, France. He is considered as an expert in sample preparation and observation using electron microscope techniques, as well as multispectral imaging tools.

Jocelyn Chanussot is Professor at the Grenoble Institute of Technology (Grenoble-INP), France, conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-

Lab). His research interests include image analysis, multicomponent image processing, nonlinear filtering, data fusion. He was visiting scholar at KTH (Sweden), NUS (Singapore), Stanford University and UCLA (USA). He is an Adjunct Professor of the University of Iceland. He is a Fellow of the IEEE and a member of the Institut Universitaire de France.

Biographies of the other authors are not available.

List of Figures

4 An example of the construction of a BPT (left). Starting from an initial partition (the leaves, here nodes 1 to 5), the most similar regions are merged, creating a new node in the tree, and a new partition. This process is iterated until only one region remains (the root of the tree). Two possible prunings of the BPT (right). We can only select nodes that form a partition of the initial image.

5 Flowchart of the proposed method. The blue boxes correspond to the segmentation part of the process, while the red ones correspond to the classification part. Green circles show the parts where the process interacts with the user, to improve the segmentation map or the training samples.

6 An example of the block matching procedure to find similar blocks in a gray-scale image.[34] A reference block (R) is shown in red, and the blue blocks (found here along an edge of a spatial structure of the image) were identified as similar.

7 The histograms of a heavily quantized band, before and after addition of uniform noise for a better variance estimation with MAD.

8 Two distances suited for the region model of (1), *i.e.* $\theta = SAM(x, y)$, and $\Delta = ||x - y||_2$, represented for a 3 dimensional space.[37]

9 The rationale behind a pruning strategy based on size discontinuity in a branch: the size of the regions often present large discontinuities when they are merged with a region corresponding to a different phase. The leftmost part shows the discontinuity in the size of the regions occurring along a branch of the tree. The height is the distance from the leaf node to the node of interest. The rightmost part show the region (a fly ash) corresponding to the node situated just before the discontinuity (see the black cross on the graph), displayed on a false color composition resulting from the first three principal components.

10 An example of a node selection that does not provide a partition of the image (left) and the application of the max-tree rule (right). The nodes of the tree have been colored: the root is in green, the leaves that have not been kept in the pruning are in purple, and the nodes kept in the pruning are in red (or orange if they are also leaves).

11 Evolution of SAM between nodes at a certain height in a branch and their siblings, and the region kept in the branch with the chosen pruning strategy, using a threshold on SAM of 0.15rad and a threshold on size of 10 pixels. On the graph, the black crosses indicate (from left to right) either the leaf node, the node kept by the proposed pruning strategy or the node kept considering only the size.

12 Some bands of the synthetic dataset used for classification evaluation. A synthetic band mimicking the BSE band (left), a normal X-ray band (middle) and a quantized X-ray band (right) are shown.

13 Segmentation results on a synthetic dataset. From left to right: Color composition (PCA), Ground truth segmentation, Results of the Mean Shift algorithm, and results of the proposed BPT pruning strategy.

14 Classification results on a synthetic dataset. The true classification map is shown on the left, while the results of SVM-MRF and the proposed approach are shown on the middle and on the right, respectively.

15 Results of the denoising on two bands (leftmost column) using a 3×3 median filter (middle column) and the BM3D algorithm (rightmost column) on small parts of two X-ray bands: Aluminium (top row) and Chlore (bottom row) of a cement sample.

16 Scatter plots of the training set of Fig. 3 on the first 3 principal components before (left) and after (right) the denoising steps.

17 Segmentation results on the real dataset. A false color composition (PCA) is shown on the left. The automatic segmentation map obtained by the proposed approach is shown on the right (Every region is represented by its mean color).

18 Classification maps of the proposed method in comparison with SVM and SVM-MRF. For the information on the classes, see Fig. 3. On the top row are displayed the training image and the SVM classifation map on the raw data (without denoising). On the bottom row we show the results of SVM-MRF (left) and of the proposed approach (right), after the integration of a dozen misclassified regions in the training set.

19 Relative variation between the accumulated mean phase percentages of each class of the image of Fig. 3 throughout the processing of a full sample.

List of Tables